

Embedded density functional theory for covalently bonded and strongly interacting subsystems

Jason D. Goodpaster,¹ Taylor A. Barnes,¹ and Thomas F. Miller, III^{*1}

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

(Dated: 12 January 2013)

Embedded density functional theory (e-DFT) is used to describe the electronic structure of strongly interacting molecular subsystems. We present a general implementation of the Exact Embedding (EE) method [J. Chem. Phys. **133**, 084103 (2010)] to calculate the large contributions of the non-additive kinetic potential (NAKP) in such applications. Potential energy curves are computed for the dissociation of $\text{Li}^+\text{-Be}$, $\text{CH}_3\text{-CF}_3$, and hydrogen-bonded water clusters, and e-DFT results obtained using the EE method are compared with those obtained using approximate kinetic energy functionals. In all cases, the EE method preserves excellent agreement with reference Kohn-Sham calculations, whereas the approximate functionals lead to qualitative failures in the calculated energies and equilibrium structures. We also demonstrate an accurate pairwise approximation to the NAKP that allows for efficient parallelization of the EE method in large systems; benchmark calculations on molecular crystals reveal ideal, size-independent scaling of wall-clock time with increasing system size.

I. INTRODUCTION

Important methodological challenges in electronic structure theory include the long-timescale simulation of *ab initio* molecular dynamics and the seamless combination of high- and low-level electronic structure methods in complex systems. Methods that exploit the intrinsic locality of molecular interactions have demonstrated encouraging progress towards these goals.^{1–18}

In particular, orbital-free embedded density functional theory (e-DFT) offers a formally exact approach to electronic structure theory in which the interactions between subsystems are evaluated in terms of their electronic densities.^{1–4} Recent work has demonstrated that constructing the embedded subsystems from individual molecules leads to a linear-scaling electronic structure approach that maps naturally onto distributed-memory parallel computers,^{14,19} and it provides a systematic framework for calculating electronic excited states in condensed phase systems.^{20,21} However, approximate treatments of the non-additive kinetic potential (NAKP) limit the accuracy of this approach in applications involving strongly interacting subsystems.²² For example, severe artifacts in the structure of liquid water, including the complete absence of a second peak in the oxygen-oxygen radial distribution function, have been predicted from existing approximations to the NAKP,¹⁹ and e-DFT applications involving covalently bonded embedded subsystems have also been shown to qualitatively fail.^{22–24} The development of improved methods to address the NAKP problem will open new doors for on-the-fly, massively parallel electronic structure calculations in general, condensed-phase systems.

In this paper, we describe progress towards the development of accurate, scalable treatments for the NAKP in e-DFT. We provide the first molecular applications of our recently developed Exact Embedding (EE) method,²⁵ demonstrating that it successfully describes the break-

ing of covalent bonds and hydrogen bonds with chemical accuracy. Additionally, we introduce and numerically demonstrate a pairwise approximation to the NAKP, which allows for the scalable implementation of the EE method in large systems. Benchmark calculations are presented for systems with up to 125 molecules, demonstrating that parallel implementation of the method enables constant system-size scaling of the wall-clock calculation time.

II. THEORY

A. Orbital-free Embedded DFT

We utilize the orbital-free e-DFT formulation of Cortona¹ and Wesolowski and coworkers.^{2,3} For the case in which the total electronic density ρ_{AB} is partitioned into two subsystems, $\rho_{AB} = \rho_A + \rho_B$, the corresponding one-electron orbitals obey the Kohn-Sham Equations with Constrained Electron Density (KSCED),³

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}[\rho_A, \rho_{AB}; \mathbf{r}] \right] \phi_i^A(\mathbf{r}) = \epsilon_i^A \phi_i^A(\mathbf{r}) \quad (1)$$

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}[\rho_B, \rho_{AB}; \mathbf{r}] \right] \phi_j^B(\mathbf{r}) = \epsilon_j^B \phi_j^B(\mathbf{r}), \quad (2)$$

where $i = 1, \dots, N^A$, $j = 1, \dots, N^B$, and N^A and N^B are the number of electrons in the respective subsystems. v_{eff} is the effective potential for the coupled one-electron equations, such that

$$v_{\text{eff}}[\rho_A, \rho_{AB}; \mathbf{r}] = v_{\text{ne}}(\mathbf{r}) + v_J[\rho_{AB}; \mathbf{r}] + v_{\text{xc}}[\rho_{AB}; \mathbf{r}] + v_{\text{nad}}[\rho_A, \rho_{AB}; \mathbf{r}], \quad (3)$$

where the N_{nuc} nuclei occupy positions $\{\mathbf{R}_i\}$,

$$v_{\text{ne}}(\mathbf{r}) = - \sum_i^{N_{\text{nuc}}} \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|}, \quad (4)$$

$$v_{\text{J}}[\rho; \mathbf{r}] = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}', \quad (5)$$

$$v_{\text{xc}}[\rho; \mathbf{r}] = \left[\frac{\delta E_{\text{xc}}[\rho]}{\delta \rho} \right](\mathbf{r}), \quad (6)$$

and $E_{\text{xc}}[\rho]$ is the exchange-correlation functional. $v_{\text{nad}}[\rho_{\text{A}}, \rho_{\text{AB}}; \mathbf{r}]$ is the potential due to the non-additive kinetic energy for non-interacting electrons, such that

$$v_{\text{nad}}[\rho_{\text{A}}, \rho_{\text{AB}}; \mathbf{r}] = \left[\frac{\delta T_{\text{s}}^{\text{nad}}[\rho_{\text{A}}, \rho_{\text{B}}]}{\delta \rho_{\text{A}}} \right](\mathbf{r}), \quad (7)$$

where $T_{\text{s}}^{\text{nad}}[\rho_{\text{A}}, \rho_{\text{B}}] \equiv T_{\text{s}}[\rho_{\text{AB}}] - T_{\text{s}}[\rho_{\text{A}}] - T_{\text{s}}[\rho_{\text{B}}]$. The subsystem densities are constructed from the corresponding KS orbitals, using $\rho_{\text{A}}(\mathbf{r}) = \sum_{i=1}^{N_{\text{A}}} |\phi_i^{\text{A}}(\mathbf{r})|^2$ and $\rho_{\text{B}}(\mathbf{r}) = \sum_{j=1}^{N_{\text{B}}} |\phi_j^{\text{B}}(\mathbf{r})|^2$. Eqs. 1-7 are easily generalized for the e-DFT description of multiple embedded subsystems.^{1,19}

Two aspects of e-DFT are worth emphasizing. Firstly, like conventional Kohn-Sham density functional theory (KS-DFT), it is a theory that is exact in principle, but practical calculations must employ approximations to the unknown exchange-correlation functional. Secondly, unlike conventional KS-DFT, the embedding formulation introduces the NAKP, $v_{\text{nad}}[\rho_{\text{A}}, \rho_{\text{AB}}; \mathbf{r}]$, since the orbitals for subsystem A are not necessarily orthogonal to those of subsystem B. Without knowledge of the exact functional for the non-interacting kinetic energy, this creates a second source of approximation in e-DFT calculations. The significance of the NAKP is system-dependent, with the most severe cases including those for which the subsystem densities greatly overlap; no approximate kinetic energy functional has been previously demonstrated to yield accurate results for embedded subsystems that are connected by covalent bonds.^{3,22,23,26,27}

B. Exact Calculations of NAKP

We have recently developed the Exact Embedding (EE) method to calculate the NAKP in the e-DFT framework.²⁵ The general method can be summarized for two embedded subsystems as follows: A Levy constrained search (LCS)²⁸ or equivalent technique is first used to determine the full set of orthogonal KS orbitals, $\{\phi_i^{\text{AB}}\}$, that correspond to the total density ρ_{AB} from the latest iteration of Eqs. 1-3. Then, from the KS orbitals $\{\phi_i^{\text{AB}}\}$, $\{\phi_i^{\text{A}}\}$, and $\{\phi_i^{\text{B}}\}$, the corresponding kinetic potentials are calculated using the exact result of King and Handy,²⁹

$$v_{T_{\text{s}}}(\mathbf{r}) = \frac{\sum_{i=1}^n (-\frac{1}{2} \phi_i(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}) - \epsilon_i \phi_i(\mathbf{r})^2)}{\rho(\mathbf{r})} + \mu, \quad (8)$$

where n is the number of occupied orbitals, ϵ_i is the KS eigenvalue corresponding to orbital ϕ_i , and μ is a constant. Finally, the NAKP needed for the next iteration of Eqs. 1-3 is calculated directly from the difference

$$v_{\text{nad}}[\rho_{\text{A}}, \rho_{\text{AB}}; \mathbf{r}] = v_{T_{\text{s}}}^{\text{AB}}(\mathbf{r}) - v_{T_{\text{s}}}^{\text{A}}(\mathbf{r}), \quad (9)$$

where the superscripts in this equation indicate the orbital set to which each kinetic potential corresponds.

Rather than explicitly performing the LCS, we use the equivalent protocol of Zhao, Morrison, and Parr (ZMP)³⁰⁻³² to obtain the exact non-interacting kinetic energy and the KS orbitals $\{\phi_i^{\text{AB}}\}$. This requires solution of the following one-electron equations

$$\left[-\frac{1}{2} \nabla^2 + v_{\text{ext}}(\mathbf{r}) + v_{\text{c}}^{\lambda}(\mathbf{r}) \right] \phi_{i,\lambda}^{\text{AB}}(\mathbf{r}) = \epsilon_{i,\lambda}^{\text{AB}} \phi_{i,\lambda}^{\text{AB}}(\mathbf{r}) \quad (10)$$

in the limit $\lambda \rightarrow \infty$, where $i = 1, \dots, (N^{\text{A}} + N^{\text{B}})$, and

$$v_{\text{c}}^{\lambda}(\mathbf{r}) = \lambda \int \frac{\rho(\mathbf{r}') - \rho_{\text{AB}}(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}'. \quad (11)$$

$v_{\text{ext}}(\mathbf{r})$ corresponds to any well-behaved external potential,^{31,32} and various choices for this potential are described in Sec. III B. In practice, Eq. 10 is solved at several large, finite values of λ , and the KS orbitals and eigenvalues, as well as the final non-interacting kinetic energy, are obtained via extrapolation.³⁰⁻³² In Sec. V, we discuss a technique to robustly implement the ZMP step for NAKP calculations in large systems.

The EE method outlined in Eqs. 8 - 11 is unique in that it allows for the formally exact calculation of the total electronic density within the e-DFT framework, using integer orbital occupancies and without approximations to the NAKP. The method was previously demonstrated for atomic systems with strongly overlapping subsystem densities,²⁵ and the current paper presents its first molecular applications. We note that several other groups have also used density inversion techniques to calculate the NAKP, assuming that the total electron density is already available from another electronic structure calculation.^{24,33,34} In particular, Visscher and coworkers have applied this approach to molecular systems with the aim of developing improved non-additive kinetic energy functionals.²⁴ Furthermore, Partition DFT has been introduced as a formally exact embedding scheme in which subsystem densities are described using partially occupied orbitals, and it has been applied to one-dimensional model systems.⁶

III. IMPLEMENTATION DETAILS

We have implemented e-DFT in the Molpro quantum chemistry package,³⁵ allowing for calculation of the NAKP with either approximate functionals or the EE method. In this section, methodological and numerical aspects of the implementation are discussed.

A. Supermolecular vs. Monomolecular Basis Sets

The atom-centered basis sets used to solve the KSCED (Eqs. 1 and 2) are implemented using two different conventions.^{22,36} In the monomolecular basis set convention, the density for each embedded subsystem is described using only the basis functions that are centered on atoms belonging to that subsystem. In the supermolecular basis set convention, the density for each embedded subsystem is described using the same basis set, which includes functions that are centered on all atoms in the system. The supermolecular basis set convention provides a closer approximation to the complete basis set limit, although it is more costly.

B. ZMP Step

In our implementation, the ZMP step of the EE method is performed by solving Eq. 10 for six large, finite values of λ . The KS orbitals $\{\phi_i^{AB}\}$ are then obtained from extrapolation of the atomic orbital coefficients for the $\{\phi_{i,\lambda}^{AB}\}$, using a third-order polynomial in λ^{-1} , and normalization of the extrapolated orbitals is enforced *a posteriori*. The KS eigenvalues $\{\epsilon_i^{AB}\}$ are similarly obtained from extrapolation of the $\{\epsilon_{i,\lambda}^{AB}\}$. $T_s[\rho_{AB}]$ is calculated analytically from the extrapolated orbital coefficients, which ensures that the total energy from the EE method is bound from below by the KS-DFT energy.

In the limit $\lambda \rightarrow \infty$, the solutions to Eq. 10 are independent of the choice of external potential $v_{\text{ext}}(\mathbf{r})$,^{30–32} although $v_{\text{ext}}(\mathbf{r})$ does affect the convergence with increasing λ . Various options were thus considered, including

$$v_{\text{ext}}(\mathbf{r}) = v_{\text{ne}}(\mathbf{r}), \quad (12)$$

$$v_{\text{ext}}(\mathbf{r}) = v_{\text{ne}}(\mathbf{r}) + \left(1 - \frac{1}{N^A + N^B}\right) v_J[\rho_{AB}; \mathbf{r}], \quad (13)$$

$$v_{\text{ext}}(\mathbf{r}) = v_{\text{ne}}(\mathbf{r}) + v_J[\rho_{AB}; \mathbf{r}] + v_{\text{xc}}[\rho_{AB}; \mathbf{r}]. \quad (14)$$

At every iteration of the KSCED, these versions of $v_{\text{ext}}(\mathbf{r})$ are all available without the need for additional computation. Test calculations have indicated that the external potential in Eq. 14 leads to the fastest convergence of the extrapolation with increasing λ , and this potential is used in all results for the EE method reported in Sec. IV.

C. NAKP Numerics for Regions of Weak Density Overlap

Numerical evaluation of the kinetic potential from Eq. 8 is unstable in regions for which the corresponding density vanishes. The problem is exacerbated by the incorrect distance dependence of the low-density tails obtained from calculations using Gaussian-type orbitals (GTOs).²⁹ However, these numerically treacherous regions correspond to weak overlap between subsystem densities, where the magnitude of the NAKP is necessarily small and easily approximated.² We thus utilize

a density-based criterion to switch from the exact expression for the kinetic potential to a numerically stable approximation, such as the Thomas-Fermi (TF) kinetic potential. The protocol used to perform this switching is described below.

In a first step, we calculate the constant shift that is needed to match the exact result for each kinetic potential to the corresponding TF result in a prescribed switching region. Specifically, for each of the kinetic potentials (i.e., $v_{T_s}(\mathbf{r}) \in \{v_{T_s}^{AB}(\mathbf{r}), v_{T_s}^A(\mathbf{r}), v_{T_s}^B(\mathbf{r})\}$ which correspond respectively to $\rho(\mathbf{r}) \in \{\rho_{AB}(\mathbf{r}), \rho_A(\mathbf{r}), \rho_B(\mathbf{r})\}$, the average difference ($\Delta \in \{\Delta^{AB}, \Delta^A, \Delta^B\}$) between the results from Eq. 8 and from the TF functional is evaluated in the vicinity of the $\rho(\mathbf{r}) = \rho'$ density isosurface. Each Δ is computed over gridpoints in the region $\xi < f[\rho; \mathbf{r}] < (1 - \xi)$, where

$$f[\rho; \mathbf{r}] = \frac{1}{e^{\kappa(\rho(\mathbf{r}) - \rho')} + 1}, \quad (15)$$

ξ , κ , and ρ' are parameters that define the switching region, and the relative contribution from each gridpoint is weighted according to

$$\omega[\rho; \mathbf{r}] = e^{-\kappa(\rho_{AB}(\mathbf{r}) - \rho(\mathbf{r}))}. \quad (16)$$

Note that the weighting function in Eq. 16 is uniform for the case of $\rho = \rho_{AB}$; for cases in which ρ is one of the subsystem densities, $\omega[\rho; \mathbf{r}]$ preferentially selects values for which $\rho(\mathbf{r}) \approx \rho_{AB}(\mathbf{r})$.

In a second step, each kinetic potential is computed on the grid; this is done by vertically shifting the exact result with the corresponding Δ and then smoothly switching to the TF result at densities below ρ' , using the density-based switching function $f[\rho; \mathbf{r}]$ in Eq. 15. Finally, the NAKP is calculated from the smoothly switched kinetic potentials using Eq. 9. The vertical shifts that are applied to kinetic potentials simply give rise to an additive constant in the final NAKP, which has no physical effect. Although we find that switching to the TF functional at low densities is both convenient and accurate, the protocol described above could be performed using any approximate kinetic energy functional.

IV. RESULTS: SMALL SYSTEMS

A. Calculation Details

In this section, e-DFT calculations are presented for the dissociation curves of $(\text{H}_2\text{O})_2$ and the covalently bound $\text{Li}^+\text{-Be}$ and $\text{CH}_3\text{-CF}_3$ molecules; standard KS-DFT calculations are included for comparison. All results are obtained using the Molpro quantum chemistry package,³⁵ with KS-DFT available in the standard version and with the e-DFT method implemented in our modified version. In the e-DFT calculations, the NAKP is described using either the EE method or the approximate TF^{37,38} and LC94³⁹ kinetic energy functionals;

these approaches will hereafter be referred to as e-DFT-EE, e-DFT-TF, and e-DFT-LC, respectively.

All calculations in this section are performed using the B88-P86 exchange-correlation (XC) functional.^{40,41} Both the XC functional and the NAKP are evaluated on a grid of Becke-Voronoi⁴² cells with resolution to limit the integration error of Slater exchange to 10^{-12} Hartree; the grid is generated using the Molpro directive GRID=10⁻¹².

The KSCED in Eqs. 1-2 are initialized from the gas phase density of each subsystem, and the eigensolutions for each set of equations are updated at every iteration. Convergence of these equations is improved with the molecular orbital (MO) shifting and direct inversion of iterative subspace (DIIS) algorithms.^{43,44} For the water dimer, an MO shift of -0.5 Hartree is employed, whereas a -1.0 Hartree shift is used for Li⁺-Be and CH₃-CF₃. Since the DIIS algorithm leads to slow final convergence,⁴⁵ it is discontinued once the root mean squared difference (RMSD) of total density matrix elements changes by less than 5×10^{-4} between two successive iterations. The KSCED equations are deemed converged when the total energy of the system changes by less than 10^{-6} Hartree and the RMSD in the total density matrix is smaller than 10^{-5} between two successive iterations.

For the ZMP step, extrapolation of the solutions to Eq. 10 is performed using $\lambda = \gamma + \tau j$, where $j = 0, 1, \dots, 5$. Unless otherwise noted, calculations for the water dimer and Li⁺-Be employ $\gamma = 5000$ and $\tau = 100$, whereas calculations for CH₃-CF₃ employ $\gamma = 100$ and $\tau = 10$. To reach adequate convergence, Eq. 10 is solved in several stages. Firstly, a coarse solution is reached by using an MO shift of -10^3 Hartree and a value of $\lambda = 100$. Subsequently, using this coarse solution as a starting point, the Eq. 10 solved using a smaller MO shift of -84 Hartree and with $\lambda = \gamma$. Finally, solution of Eq. 10 for each increasing value of λ needed for extrapolation employs the solution for the prior value of λ as a starting point. The DIIS algorithm is used throughout. The orbitals from Eq. 10 are deemed converged when the RMSD in the density matrix was smaller than 10^{-9} between two successive iterations; significantly tighter convergence is needed for the ZMP equations than for the KSCED, to ensure an accurate extrapolation.

Calculations for the water dimer variously employ the aug-pc-3, aug-pc-2, and aug-pc-1 basis sets,⁴⁶ in each case using only the s- and p-type functions for the hydrogen atoms and the s-, p-, and d-type functions for the oxygen atoms. These water dimer basis sets are hereafter referred to as the modified aug-pc-3, aug-pc-2, and aug-pc-1 basis sets, respectively. Calculations involving Li⁺-Be use the s-, p-, and d-type functions of the combined aug-pc-4 and cc-pVQZ (core/valence) basis sets.⁴⁷ In calculations for CH₃-CF₃, the C atoms are described using the s-, p-, and d-type functions of the combined aug-pc-4 and cc-pV6Z (core/valence) basis sets,⁴⁷ and the H and F atoms are described using the full aug-pc-1 basis set.⁴⁶ Sensitivity of the e-DFT calculations to the

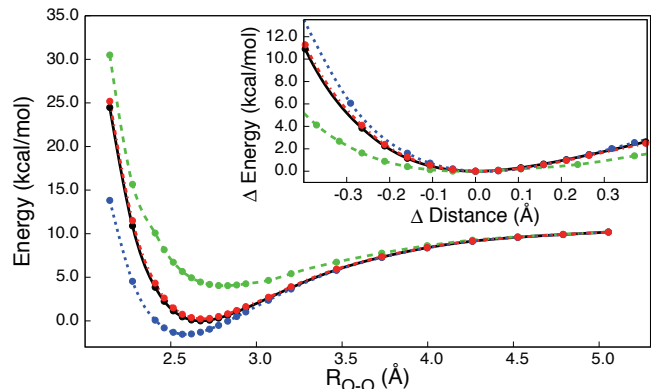


FIG. 1. The water dimer dissociation curve, obtained using e-DFT-EE (red, dot-dashed), e-DFT-TF (green, dashed) and e-DFT-LC (blue, dotted). Also included are reference KS-DFT results (black, solid), which are graphically indistinguishable from the e-DFT-EE results. Total energies are plotted with respect to the KS-DFT minimum of -152.430722 Hartree. Inset, the curves are shifted vertically to align the energy minima and horizontally to align the equilibrium distances.

basis set is discussed in the next section.

Larger basis sets provide a better description of low-density regions, allowing for the use of smaller values for the parameter ρ' in Eqs. 15 and 16 and providing robustness with respect to the choice of this parameter. For the water dimer, calculations using aug-pc-3, aug-pc-2, and aug-pc-1 basis sets employ values of $\rho' = 10^{-5}$, 10^{-4} , and 5×10^{-3} , respectively. For Li⁺-Be and CH₃-CF₃, calculations employ $\rho' = 10^{-6}$. In each case, $\xi = 10^{-4}$, and the parameter κ in Eqs. 15 and 16 is chosen such that $\kappa\rho' = 10$.

B. Water Dimer

Fig. 1 presents the dissociation curve for the water dimer, a system with a strong hydrogen bond and significantly overlapping subsystem densities. The curve is obtained using e-DFT-EE (dot-dashed), e-DFT-TF (dashed), and e-DFT-LC (dotted); KS-DFT results (solid) are also included for reference. The e-DFT calculations are performed using two embedded subsystems, each corresponding to a different molecule in the dimer. All calculations presented in the figure utilize the modified aug-pc-3 basis set, with the e-DFT calculations employing the supermolecular basis set convention. The dissociation curve is plotted as a function of the oxygen-oxygen distance, with the equilibrium water dimer geometry obtained from a KS-DFT energy minimization and with other geometries obtained by displacing the two molecules along the oxygen-oxygen vector while fixing all other internal coordinates.

The e-DFT-EE results in Fig. 1 agree well with KS-DFT throughout the range of dissociation distances. Numerical results for the two methods are graphically indis-

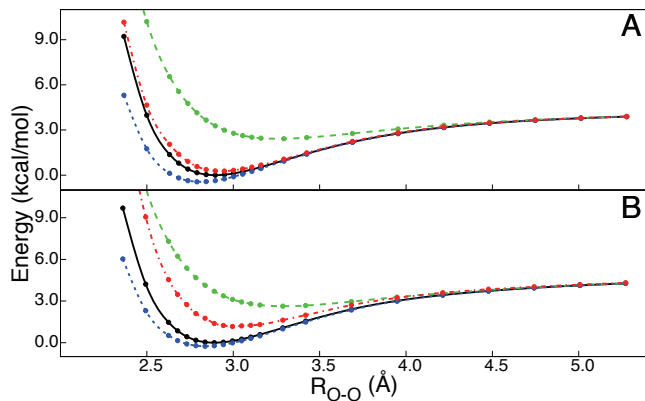


FIG. 2. Basis set dependence of the water dimer dissociation curve, illustrated for calculations using the (A) modified aug-pc-2 and (B) modified aug-pc-1 basis sets. Results for the e-DFT-EE, e-DFT-TF, e-DFT-LC, and KS-DFT methods are reported as in Fig. 1. Total energies are plotted with respect to the KS-DFT minimum energies of -152.953947 Hartree (panel A) and -152.864441 Hartree (panel B).

tinguishable, and the calculated total energies differ by less than 0.5 kcal/mol throughout the entire attractive branch of the curve. Exact numerical agreement between the e-DFT-EE and KS-DFT descriptions is expected only in the limits of an exact XC functional and a complete basis set.

The sensitivity of the e-DFT results to approximations in the NAKP is clearly demonstrated in Fig. 1. The curve obtained using e-DFT-TF differs significantly from the KS-DFT reference, exhibiting a dissociation energy that is underestimated by 40% (~ 4 kcal/mol) and an equilibrium bond length that is 0.15 Å too long. Calculations obtained using e-DFT-LC are somewhat improved, although the dissociation energy is still overestimated by 20% (~ 2 kcal/mol) and the equilibrium bond length is underestimated by 0.10 Å. In the inset of Fig. 1, the curvature of the potential energy surfaces in the vicinity of the minimum are compared, revealing significant deviations of the results obtained using the approximate NAKP treatments (e-DFT-TF and e-DFT-LC) with respect to the results obtained using KS-DFT and e-DFT-EE.

Iannuzzi and coworkers¹⁹ have demonstrated that e-DFT calculations using approximate treatments of the NAKP, including the TF and LC94 functionals, lead to qualitative failure in describing the structure of liquid water. Fig. 1 illustrates the origin of this failure in terms of the pairwise interactions among molecules, and it suggests that e-DFT-EE will enable the accurate, first-principles simulation of liquid water and aqueous solutions. Critical to this effort, however, is the efficient and parallelizable implementation of the EE method for large systems, which is discussed in Section V.

The sensitivity of the e-DFT calculations to basis set completeness is illustrated in Fig. 2, in which the water dimer dissociation curves are recalculated using the modified aug-pc-2 (Fig. 2A) and modified aug-pc-1 basis sets

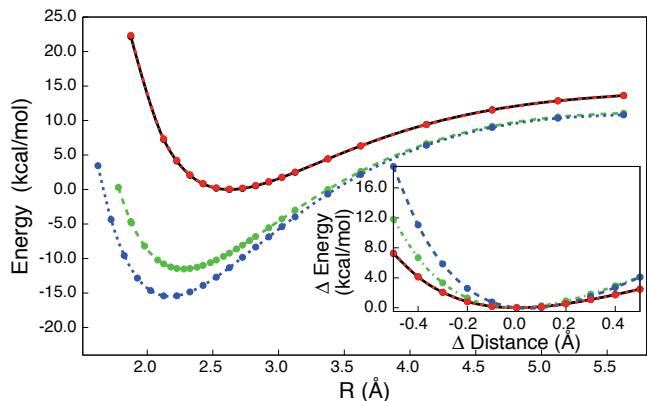


FIG. 3. The $\text{Li}^+\text{-Be}$ dissociation curve. Results for the e-DFT-EE, e-DFT-TF, e-DFT-LC, and KS-DFT methods are reported as in Fig. 1. The results for e-DFT-EE and the reference KS-DFT results are graphically indistinguishable. Total energies are plotted with respect to the KS-DFT minimum energy of -21.962072 Hartree. Inset, the curves are aligned as in the inset of Fig. 1.

(Fig. 2B). Comparison of the KS-DFT results and the e-DFT-EE results reveals that the agreement between the methods worsens with smaller basis set; of course, both the KS-DFT calculations and the e-DFT-EE calculations are basis-set dependent. In the e-DFT-EE calculations, smaller basis sets give rise to numerical artifacts including the oscillatory behavior in the King-Handy expression for the kinetic potential.²⁹ For the modified aug-pc-1 basis set (Fig. 2B), the reasonable agreement between KS-DFT and e-DFT-LC is due to a fortuitous cancellation of errors from the approximate NAKP functional and the small basis set.

C. $\text{Li}^+\text{-Be}$

We now consider the heterolytic cleavage of a weak covalent bond, $\text{Li}^+\text{-Be} \rightarrow \text{Li}^+ + \text{Be}$, using KS-DFT and e-DFT. The e-DFT calculations were performed in the supermolecular basis set convention using two embedded subsystems, one corresponding to the 2-electron Li ion and the other corresponding to the 4-electron Be atom. The dissociation curve for $\text{Li}^+\text{-Be}$ is plotted in Fig. 3.

As is seen from the main figure, the e-DFT-EE calculations accurately reproduce the total energies from KS-DFT throughout the entire range of internuclear distances. The dissociation curves for these two methods, which are graphically indistinguishable in Fig. 3, deviate by less than 0.2 kcal/mol throughout the range of separations and the dissociation energy deviates by only 0.07 kcal/mol. In contrast, the e-DFT-TF results are in qualitative disagreement with the KS-DFT reference calculations; in addition to dramatically overestimating the dissociation energy of the molecule by approximately 12.5 kcal/mol, the method predicts the equilibrium bond length to be 20% too short. Interestingly, the e-DFT-LC method performs significantly worse in this application.

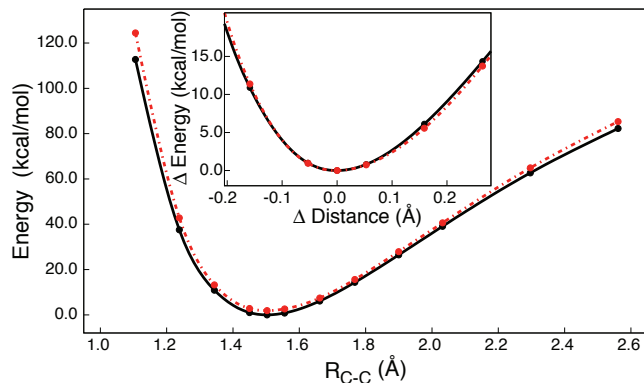


FIG. 4. The $\text{CH}_3\text{-CF}_3$ dissociation curve for heterolytic cleavage of the C-C bond. Results are presented for the e-DFT-EE (red, dot-dashed) and KS-DFT (black, solid) methods. Total energies are plotted with respect to the KS-DFT minimum energy of -377.575687 Hartree. Inset, the curves are aligned as in the inset of Fig. 1.

The calculations based on the approximate LC94 kinetic energy functional overestimate the dissociation energy by approximately 16 kcal/mol and predict the equilibrium bond length to be 25% too short. The inset to Fig. 3 illustrates that both e-DFT methods that use approximate treatments for the NAKP lead to an overestimation of the energy surface curvature in the vicinity of the equilibrium bond distance.

The results in Fig. 3 illustrate the well-known breakdown of e-DFT with approximate treatments of the NAKP for applications involving strongly overlapping subsystem densities. They further show that our EE method overcomes this large error, yielding the first numerical demonstration of an e-DFT method to describe covalent bond-breaking with chemical accuracy. Since e-DFT-EE is a formally exact method, this result is expected. However demonstration that the level of accuracy in Fig. 3 can be achieved in practical numerical simulations constitutes a non-trivial validation of the method.

D. $\text{CH}_3\text{-CF}_3$

In a more challenging application for e-DFT, we consider the heterolytic cleavage of a strong carbon-carbon σ -bond, $\text{CH}_3\text{-CF}_3 \rightarrow \text{CH}_3^+ + \text{CF}_3^-$. The e-DFT calculations were again performed in the supermolecular basis set convention using two embedded subsystems, one corresponding to the 8-electron CH_3^+ moiety and the other corresponding to the 34-electron CF_3^- moiety. The geometry for the lowest energy point along the curve is provided in the supplemental information; the dissociation curve in Fig. 4 is plotted by extending the C-C distance while keeping all other internal coordinates unchanged.

The dissociation curves in Fig. 4 are presented only for e-DFT-EE and the reference KS-DFT calculations. e-DFT-EE reproduces the KS-DFT reference value for the total energy for the molecule at the equilibrium bond dis-

tance to within 1.5 kcal/mol, and the embedding method also recovers the reference value for the equilibrium bond distance. Furthermore, as is clear from the inset, e-DFT-EE accurately reproduces the curvature of the energy surface in the vicinity of the equilibrium bond distance. In contrast, the e-DFT-TF and e-DFT-LC descriptions for this system fail dramatically, predicting total energies at the equilibrium bond distance that deviate from the KS-DFT reference by approximately 730 kcal/mol and 980 kcal/mol, respectively. For calculations with such strongly interacting subsystems, the failure of e-DFT with approximate descriptions for the NAKP methods has been previously observed.²² However, the results for e-DFT-EE in Fig. 4 demonstrate significant progress in the accurate description of covalently interacting subsystems using e-DFT.

V. RESULTS: EXTENSION TO LARGER SYSTEMS

A. Pairwise treatment of the NAKP

In the previously described implementation of e-DFT-EE, the ZMP step, or an equivalent LCS, is performed on the full system of interest. However, numerical challenges limit the LCS to systems with less than 10-15 atoms,^{33,34,48-51} potentially hindering the applicability of e-DFT-EE in large systems. To avoid this problem, we demonstrate a pairwise approximation for the NAKP that enables the scalable implementation of e-DFT-EE.

For a system composed of N_{sub} embedded subsystems, $\{\rho_\alpha\}$, the non-additive kinetic energy can be approximated using a pairwise sum,²⁵ such that

$$T_s^{\text{nad}}[\{\rho_\alpha\}] \equiv T_s[\bar{\rho}] - \sum_{\alpha=1}^{N_{\text{sub}}} T_s[\rho_\alpha] \quad (17)$$

$$\approx \sum_{\alpha < \beta = 1}^{N_{\text{sub}}} (T_s[\rho_\alpha + \rho_\beta] - T_s[\rho_\alpha] - T_s[\rho_\beta]),$$

where $\bar{\rho} = \sum_{\alpha=1}^{N_{\text{sub}}} \rho_\alpha$. The NAKP for a given subsystem α is then

$$v_{\text{nad}}[\rho_\alpha, \{\rho_\alpha\}; \mathbf{r}] = \sum_{\beta \neq \alpha}^{N_{\text{sub}}} (v_{T_s}^{\alpha\beta}(\mathbf{r}) - v_{T_s}^\alpha(\mathbf{r})). \quad (18)$$

Applying the EE method to this approximation for the NAKP, a ZMP step is performed at each iteration of the KSCED to obtain the KS orbitals corresponding to each pair of subsystems densities, $\{\phi_i^{\alpha\beta}\}$. Then, using both the subsystem KS orbitals $\{\phi_i^\alpha\}$ from the KSCED and the subsystem-pair KS orbitals $\{\phi_i^{\alpha\beta}\}$, the NAKP is evaluated directly from Eqs. 8 and 18. In this approach, only the NAKP is assumed to be pairwise additive; all other interactions in the system are treated with full generality. Since the ZMP step is applied only to the subsystem

pairs, this approach is numerically feasible if each subsystem is limited to a relatively small number of atoms, regardless of the total system size. The short-ranged nature of contributions to the non-additive kinetic energy suggests that distance-based cutoffs can be employed within the sum over subsystem pairs.²⁵

It was emphasized earlier that the converged results of the ZMP step are independent of the choice of external potential, $v_{\text{ext}}(\mathbf{r})$, in Eq. 10. In the pairwise implementation of e-DFT-EE for the water trimer in Sec. V B, we employ the following external potential for each pair of densities ρ_α and ρ_β ,

$$v_{\text{ext}}(\mathbf{r}) = v_{\text{ne}}(\mathbf{r}) + v_{\text{J}}[\bar{\rho}; \mathbf{r}] + v_{\text{xc}}[\bar{\rho}; \mathbf{r}] + \frac{\delta \tilde{T}_{\text{s}}[\bar{\rho}]}{\delta(\rho_\alpha + \rho_\beta)} - \frac{\delta \tilde{T}_{\text{s}}[\rho_\alpha + \rho_\beta]}{\delta(\rho_\alpha + \rho_\beta)}, \quad (19)$$

where \tilde{T}_{s} indicates the approximate TF functional. This external potential approximates the KSCED effective potential (Eq. 3) for the pair of subsystems embedded within the remainder of the full system; note that the TF functional is used only to regularize the effective potential for the ZMP step; it does not introduce any additional approximation into the e-DFT-EE calculation. In Sec. V C, we use a simple external potential that includes only the electron-nuclear interactions for the subsystem pair.

The following two sections demonstrate the accuracy of this pairwise implementation of e-DFT-EE (Sec. V B) and the efficiency with which it can be implemented in parallel (Sec. V C).

B. Water Trimer Application: Testing Pairwise Additivity in the NAKP

Fig. 5 presents a test of pairwise additivity in the NAKP (Eq. 18) for a hydrogen-bonded trimer of water molecules. e-DFT-EE calculations are performed using three embedded subsystems, each corresponding to a different molecule in the trimer. In a first set of results, the symmetric dissociation curve for the trimer is calculated using no assumptions about the NAKP (solid); in a second set of results, the curve is calculated using the assumption of pairwise additivity of the NAKP (dot-dashed). The equilibrium geometry is provided in the supplemental information; other geometries along the dissociation curve were then obtained by uniformly stretching the oxygen-oxygen distances in the cluster, keeping all other internal coordinates unchanged. The trimer calculations were performed using the modified aug-pc-2 basis set with the monomolecular basis set convention; all other calculation details are identical to those described previously for the modified aug-pc-2 calculations of the water dimer.

The agreement between the two curves in Fig. 5 indicates that Eqs. 17 and 18 are excellent approximations for the non-additive kinetic energy and NAKP, respectively. Throughout the entire attractive branch of the

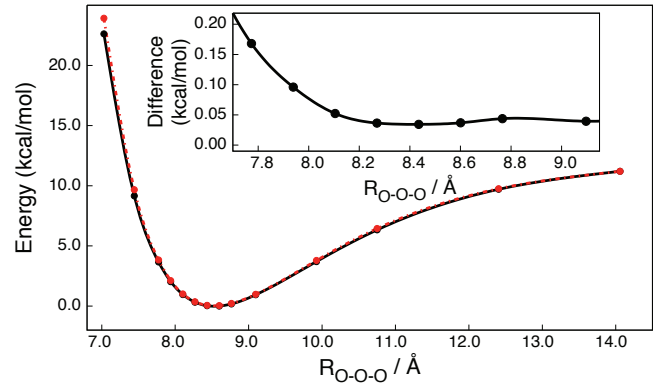


FIG. 5. Symmetric dissociation curves for the water trimer, illustrating the pairwise additivity of the NAKP. Calculations are performed using the e-DFT-EE method, with no approximation to the NAKP (black, solid) and with the pairwise approximation to NAKP (red, dot-dashed). The curves are plotted as a function of the sum of the three O-O distances, with details of the molecular geometries provided in the text. Total energies plotted with respect to the minimum energy of -229.440307 Hartree for the full NAKP treatment. Inset, the difference between the two curves is plotted.

curve the total energies differ by less than 0.5 kcal/mol, and the largest deviations appear only in the strongly repulsive region at short distances. This good agreement is particularly notable, given that the cyclic trimer geometries might be expected to magnify possible non-additive contributions to the total energy; even better adherence of the NAKP to pairwise additivity is expected for linear geometries of the trimer. We have previously noted that higher-order corrections to Eqs. 17 and 18 are possible,²⁵ although the results in Fig. 5 suggest that the assumption of pairwise additivity will be adequate in many cases.

C. Parallel Scaling of e-DFT-EE

Primary bottlenecks in KS-DFT include calculation of the two-electron integrals and solution of the eigenvalue problem. In standard implementations, the two-electron integral calculations scale as M^4 and the eigenvalue calculation scales at best as M^2 , where M is the total number of basis functions.^{52,53} More efficient methods for computing the two-electron integrals include prescreening,⁵⁴ Ewald summations,⁵⁵ and the fast-multipole method;⁵⁶ however, solution of the eigenvalue problem remains a computational bottleneck in most KS-DFT implementations.⁵⁷

As has been noted in previous work,¹⁹ the monomolecular basis set convention leads to advantageous scaling properties for e-DFT. In this convention, the number of basis functions used to solve each KSCED, M_{sub} , is independent of system size. Consequently, the total cost of the eigenvalue problem scales linearly with the number of subsystems, N_{sub} , and it can be trivially parallelized to the subsystem level.

The cost of the two-electron integral calculation is also reduced in the monomolecular basis set convention. Terms arising from orbitals centered on molecules in more than two different subsystems are exactly zero, such that the total cost of this operation scales with $N_{\text{sub}}^2 M_{\text{sub}}^4$. Furthermore, in this convention, the density for each subsystem is spatially localized, such that short-ranged contributions to the KSCED effective potential, including exchange, correlation, short-ranged electrostatic contributions, and pair-wise contributions to the NAKP, can be truncated at a cutoff distance. Long-ranged electrostatic contributions to the KSCED effective potential can be efficiently treated using Ewald summations or other standard methods.^{55,56} Setting aside these long-ranged terms for the current demonstration, the use of distance-based cutoffs reduces the scaling of the total two-electron integral calculation to $N_{\text{sub}} M_{\text{sub}}^4$, which can be parallelized to yield constant wall-clock time scaling with increasing system size.

To illustrate these scaling properties, Fig. 6 presents benchmark timings for simple tetragonal lattices of 8 to 125 H_2 molecules, using both e-DFT-EE and the KS-DFT implementation in Molpro. The H_2 molecules are oriented parallel to the z axis, with a bond length of 0.8 Å, and the centers-of-mass for the molecules are spaced by 3.0 Å along the x and y axes and by 3.8 Å along the z axis. All calculations employ the uncontracted STO-3G basis set,⁵⁸ Slater exchange⁵⁹ without electron correlation, and a grid density that ensures an integration error in the exchange energy of less than 10^{-6} Hartree. The e-DFT-EE calculations are performed with each molecule defined as a different subsystem, using the monomolecular basis set convention, and using one parallel processor per subsystem. Values for the parameters λ , ρ' , κ , and the MO shift are the same as those used for the $\text{Li}^+\text{-Be}$ system. The cutoff for the calculation of the electrostatics, exchange, and NAKP terms is set to 4.0 Å in these calculations, such that only nearest-neighbor molecules in the lattice contribute to these terms. All calculations are performed on a cluster of dual, quad-core 2.6 GHz Xeon Intel processors with Infiniband communication.

The timings in Fig. 6 indicate that the e-DFT-EE wall-clock time scales independently of the system size, with the deviations at small sizes due to the boundaries of the finite crystal. As expected, the KS-DFT results in the serial Molpro implementation with integral pre-screening scales quadratically with the increasing system size. In Fig. 7, relative energy of the e-DFT-EE and the KS-DFT calculations are plotted as a function of the number nearest-neighbor pairs in the lattice, $N_{\text{pairs}} = 3(N_{\text{sub}} - N_{\text{sub}}^{2/3})$. The error is small and independent of system size. The integrated error in the density per molecule was found to behave similarly (not shown).

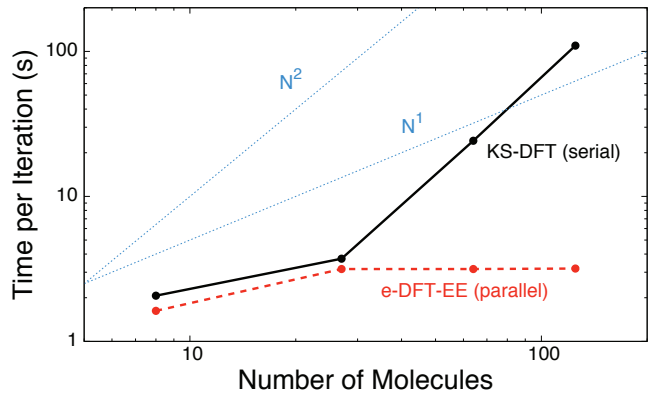


FIG. 6. Wall-clock timings for lattices of hydrogen molecules, ranging in size from 8 to 125 H_2 molecules. The dotted blue lines indicate ideal quadratic and linear scaling, the solid, black curve corresponds to the serial implementation of integral-prescreened KS-DFT in Molpro, and the dashed, red curve corresponds to e-DFT-EE using a number of parallel processors equal to the number of molecules in the system.

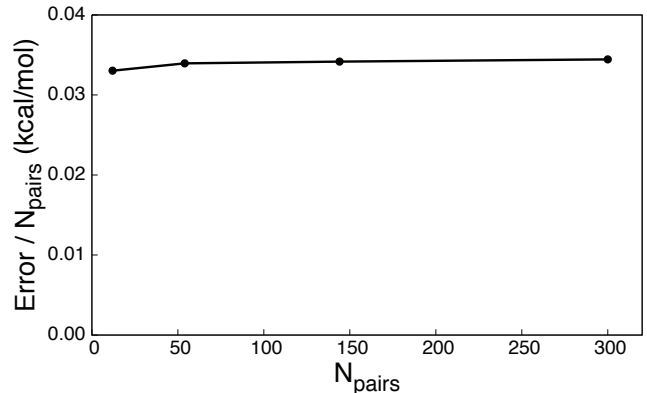


FIG. 7. Error in the total energy of the e-DFT-EE calculation relative to KS-DFT for increasing system size, plotted with respect to the number of nearest-neighbor pairs.

CONCLUSIONS

We introduce a general implementation of the EE method for calculating NAKP contributions in the e-DFT framework, and we present a range of molecular applications. The accuracy of e-DFT-EE is demonstrated for systems with covalently bonded and hydrogen-bonded subsystems. For the dissociation of the water dimer and the covalent bonds in $\text{Li}^+\text{-Be}$ and $\text{CH}_3\text{-CF}_3$, e-DFT-EE preserves excellent agreement with reference KS-DFT calculations, whereas approximate treatments for the NAKP, including those based on the TF or LC94 kinetic energy functionals, lead to known failures. Furthermore, pairwise approximation of the NAKP yields excellent accuracy for the hydrogen-bonded water trimer, and it enables ideal, constant system-size scaling in applications to molecular clusters with up to hundreds of

atoms. These results establish e-DFT-EE as a promising methodology for performing accurate, first-principles molecular dynamics and for accurately embedding high-level wavefunction methods in complex systems.

ACKNOWLEDGEMENTS

This work is supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant W911NF-10-1-0202 and by the U. S. Office of Naval Research under grant N00014-10-1-0884. T.A.B. acknowledges support from a National Defense Science and Engineering Graduate Fellowship, and T.F.M. acknowledges support from a Camille and Henry Dreyfus Foundation New Faculty Award and an Alfred P. Sloan Foundation Research Fellowship.

- ¹P. Cortona, Phys. Rev. B **44**, 8454 (1991).
- ²T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).
- ³T. A. Wesolowski, in *Computational Chemistry: Reviews of Current Trends - Vol. 10*, pp. 1-82 (World Scientific, Singapore, 2006).
- ⁴N. Govind, Y. A. Wang, A. J. R. da Silva, and E. A. Carter, Chem. Phys. Lett. **295**, 129 (1998).
- ⁵H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
- ⁶P. Elliott, K. Burke, M. H. Cohen, and A. Wasserman, Phys. Rev. A **82**, 024501 (2010).
- ⁷W. T. Yang, Phys. Rev. A **44**, 7823 (1991).
- ⁸G. K.-L. Chan, J. Chem. Phys. **120**, 3172 (2004).
- ⁹W. Z. Liang, C. Saravanan, Y. H. Shao, R. Baer, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. **119**, 4117 (2003).
- ¹⁰S. Li, J. Shen, W. Li, and Y. Jiang, J. Chem. Phys. **125**, 074109 (2006).
- ¹¹W. Li, P. Piecuch, J. R. Gour, and S. Li, J. Chem. Phys. **131**, 114109 (2009).
- ¹²D. G. Federov and K. Kitaura, J. Chem. Phys. A **111**, 6904 (2007).
- ¹³M. A. Collins and V. A. Deev, J. Chem. Phys. **125**, 104104 (2006).
- ¹⁴C. R. Jacob, J. Neugebauer, and L. Visscher, J. Comput. Chem. **29**, 1011 (2008).
- ¹⁵J. Sauer, Chem. Rev. **89**, 199 (1989).
- ¹⁶G. Paccioni, P. S. Pagus, and F. Parmigiani, eds. *Cluster models for surface and bulk phenomena*, pp. 1-82 (NATO ASI Ser., New York, Plenum, 1992).
- ¹⁷K. Jug and T. Bredow, J. Comput. Chem. **25**, 1551 (2004).
- ¹⁸D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).
- ¹⁹M. Iannuzzi, B. Kirchner, and J. Hutter, Chem. Phys. Lett. **421**, 16 (2006).
- ²⁰J. W. Kaminski, S. Gusarov, T. A. Wesolowski, and A. Kovalenko, J. Phys. Chem. A **114**, 6082 (2010).
- ²¹A. S. P. Gomes, C. R. Jacob, and L. Visscher, Phys. Chem. Chem. Phys. **10**, 5353 (2008).
- ²²A. W. Götz, S. M. Beyhan, and L. Visscher, J. Chem. Theory Comput. **5**, 3161 (2009).
- ²³S. Fux, K. Kiewish, C. R. Jacob, J. Neugebauer, and M. Reiher, Chem. Phys. Lett. **461**, 353 (2008).
- ²⁴S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).
- ²⁵J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller, III, J. Chem. Phys. **133**, 084103 (2010).
- ²⁶C. R. Jacob and L. Visscher, J. Phys. Chem. **128**, 155102 (2008).
- ²⁷S. M. Beyhan, A. W. Götz, C. R. Jacob, and L. Visscher, J. Chem. Phys. **132**, 044114 (2010).
- ²⁸M. Levy, Proc. Natl. Acad. Sci. USA **76**, 6062 (1979).
- ²⁹R. A. King and N. C. Handy, Phys. Chem. Chem. Phys. **2**, 5049 (2000).
- ³⁰Q. S. Zhao and R. G. Parr, Phys. Rev. A **46**, 2337 (1992).
- ³¹Q. S. Zhao and R. G. Parr, J. Chem. Phys. **98**, 543 (1993).
- ³²Q. S. Zhao, R. C. Morrison, and R. G. Parr, J. Chem. Phys. **50**, 2138 (1994).
- ³³O. Roncero, M. P. de Lara-Castells, P. Villarreal, F. Flores, J. Ortega, M. Paniagua, and A. Aguado, J. Chem. Phys. **129**, 184104 (2008).
- ³⁴O. Roncero, A. Zanchet, P. Villarreal, and A. Aguado, Chem. Phys. Lett. **131**, 234110 (2009).
- ³⁵MOLPRO, version 2008.3, a package of ab initio programs, H.-J. Werner, P. J. Knowles, F. R. Manby, M. Schütz, and others, see www.molpro.net.
- ³⁶T. A. Wesolowski, J. Chem. Phys. **106**, 8516 (1997).
- ³⁷L. H. Thomas, Proc. Camb. Phil. Soc. **23**, 542 (1927).
- ³⁸E. Fermi, Z. Phys. **48**, 73 (1928).
- ³⁹A. Lembarki and H. Chermette, Phys. Rev. A **50**, 5328 (1994).
- ⁴⁰J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).
- ⁴¹A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
- ⁴²C. W. Murray, N. C. Handy, and G. J. Laming, Mol. Phys. **78**, 997 (1993).
- ⁴³P. Pulay, Chem. Phys. Lett. **72**, 393 (1980).
- ⁴⁴P. Pulay, J. Chem. Phys. **3**, 556 (1982).
- ⁴⁵D. M. Shaw and A. St-Amant, J. Theor. Comput. Chem. **3**, 419 (2004).
- ⁴⁶F. Jensen, J. Chem. Phys. **115**, 9113 (2001).
- ⁴⁷D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **103**, 4572 (1995).
- ⁴⁸F. A. Hamprecht A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).
- ⁴⁹A. D. Boese N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**, 1670 (2000).
- ⁵⁰P. J. Wilson T. J. Bradley, and D. J. Tozer, J. Chem. Phys. **115**, 9233 (2001).
- ⁵¹G. Menconi P. J. Wilson, and D. J. Tozer, J. Chem. Phys. **114**, 3958 (2001).
- ⁵²S. Delvaux and M. Van Barel, J. Comput. Appl. Math. **213**, 268 (2008).
- ⁵³M. Gu and S. C. Eisenstat, SIAM J. Matrix Anal. Appl. **16**, 172 (1995).
- ⁵⁴M. Schutz, R. Lindh, and H.-J. Werner, Molecular Phys. **96**, 719 (1999).
- ⁵⁵S. Hammes-Schiffer and H. C. Andersen, J. Chem. Phys. **101**, 375 (1994).
- ⁵⁶G. Te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. Van Gisbergen, J. G. Snijders, and T. Ziegler, J. Comput. Chem. **22**, 931 (2001).
- ⁵⁷Y. Saad, J. R. Chelikowsky, and S. M. Shontz, SIAM Review **52**, 3 (2010).
- ⁵⁸W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
- ⁵⁹J. C. Slater, Phys. Rev. **81**, 385-390 (1951).